



**UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office**

Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

Cu

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.
-----------------	-------------	----------------------	---------------------

09/100,189 06/18/98 NUNBERG G D/95465

LM12/0313

RONALD ZIBELLI
XEROX CORPORATION
XEROX SQUARE 20A
ROCHESTER NY 14644

EXAMINER

THOMAS, J

ART UNIT

PAPER NUMBER

2747

DATE MAILED:

03/13/00

Please find below and/or attached an Office communication concerning this application or proceeding.

Commissioner of Patents and Trademarks

Office Action Summary

Application No.
09/100,189

Applicant(s)

NUNBERG, et al.

Examiner

JOSEPH THOMAS

Group Art Unit

2747



☐ Responsive to communication(s) filed on _____

☐ This action is **FINAL**.

☐ Since this application is in condition for allowance except for formal matters, **prosecution as to the merits is closed** in accordance with the practice under *Ex parte Quayle*, 35 C.D. 11; 453 O.G. 213.

A shortened statutory period for response to this action is set to expire 3 (THREE) month(s), or thirty days, whichever is longer, from the mailing date of this communication. Failure to respond within the period for response will cause the application to become abandoned. (35 U.S.C. § 133). Extensions of time may be obtained under the provisions of 37 CFR 1.136(a).

Disposition of Claim

☒ Claim(s) 1-27 ~~are~~ are pending in the applicat

Of the above, claim(s) _____ is/are withdrawn from consideration

☐ Claim(s) _____ is/are allowed.

☒ Claim(s) 1-27 ~~are~~ are rejected.

☐ Claim(s) _____ is/are objected to.

☐ Claims _____ are subject to restriction or election requirement.

Application Papers

☒ See the attached Notice of Draftsperson's Patent Drawing Review, PTO-948.

☐ The drawing(s) filed on _____ is/are objected to by the Examiner.

☐ The proposed drawing correction, filed on _____ is ☐ approved ☐ disapproved.

☐ The specification is objected to by the Examiner.

☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. § 119

☐ Acknowledgement is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d).

☐ All ☐ Some* ☒ None of the CERTIFIED copies of the priority documents have been

☐ received.

☐ received in Application No. (Series Code/Serial Number) _____.

☐ received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

*Certified copies not received: _____

☒ Acknowledgement is made of a claim for domestic priority under 35 U.S.C. § 119(e).

Attachment(s)

☒ Notice of References Cited, PTO-892

☒ Information Disclosure Statement(s), PTO-1449, Paper No(s). 4

☐ Interview Summary, PTO-413

☒ Notice of Draftsperson's Patent Drawing Review, PTO-948

☐ Notice of Informal Patent Application, PTO-152

— SEE OFFICE ACTION ON THE FOLLOWING PAGES —

DETAILED ACTION

Claim Objections

1. Claim 22 is objected to because of the following informalities:

The word "fact" at line 3 of claim 22 should apparently be -
-facet--. Appropriate correction is required.

Claim Rejections - 35 USC § 112

2. The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

3. Claim 5 is rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which Applicant regards as the invention.

(A) Claim 5 recites the limitation "the string recognizable constructional cue" in line 1. There is insufficient antecedent basis for this limitation in the claim itself, or in claim 1, from which it directly depends.

Claim Rejections - 35 USC § 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

5. Claims 1-3, 6, 13-14, and 26 are rejected under 35 U.S.C. 103(a) as being unpatentable over Martino, et al. (5,913,185).

(A) As per claim 1, Martino discloses a computer-implemented method of determining the language or genre of a computer recorded document and the point of shift of languages or genres therein, without using involved and time-consuming natural language and speech characteristics and analyses (Martino; abstract; col. 1, lines 25-63; and col. 3, lines 5-21), comprising the steps of:

(i) generating a log of punctuation (such as periods, commas, or quotation marks), capitalization, font, or paragraphs (read on "set of nonstructural, surface cues"), since changes therein typically signify (i.e., "cue") a language/genre change in the text (Martino; col. 10, lines 46 to col. 11, line 15 and

fig. 4); and

(ii) determining the language or genre of the text by examining the log files of punctuation, capitalization, font changes or paragraph changes to pinpoint shift points, and utilizing weighted vectors to distinguish between a plurality of genres (Martino; col. 10, line 62 to col. 11, line 15; col. 12, lines 54-61; col. 13, line 49 to col. 14, line 4; and col. 18, lines 29-38).

Martino fails to expressly teach the use of vectors, rather than log files, to represent the occurrences of punctuation, capitalization, font changes or paragraph changes. However, Martino does disclose the use of bit vectors to mathematically represent the length of a measured interval that reliably separates languages and/or genres (Martino; col. 8, lines 14-45). As such, the use of vectors to mathematically represent features obtained from a computer-recorded textual document is old and well known in the art.

One having ordinary skill in the art at the time of the invention would have found it an obvious modification to utilize vectors to mathematically represent the number of instances of changes in (or different) punctuation, capitalization, font, or paragraph, with the motivation of providing a more quantitative representation that is more succinct and requires less memory than a log file (Martino; col. 2, lines 8-9).

(B) As per claims 2-3, note Martino's teachings with respect to generating a log of punctuation including periods, commas, or quotation marks (read on "set of nonstructural, surface cues"), since changes therein typically signify (i.e., "cue") a language/genre change in the text (Martino; col. 10, lines 46 to col. 11, line 15 and fig. 4). Insofar as Applicant claim recites "represents a one of", it is irrelevant whether or not Martino also discloses punctuation cues such as dashes, question marks, and/or semi-colons, *per se*.

(C) As per claim 6, Martino fails to expressly disclose a formulae cue, *per se*. However, it is well known in the art that a formula typically involves the use of Greek or Latin variables, and since Martino clearly discloses that his invention is applicable to any language having human-readable printed symbols without regard to alphabet or word form (Martino; col. 3, lines 22-30), it would have required no hindsight for the skilled artisan to utilize a formulae cue within the Martino system by which to identify a language or genre, motivated by the fact that the use of formulae having Greek or Latin variables is more likely to appear in scientific or technical writing than in other writing genres (Martino; col. 3, lines 9-13).

(D) As per claims 13-14, note Martino's teaching with respect to

varying the length of interval analyzed for language/genre separation. In particular, Martino discloses a moving interval of between 8 and 32 word per text sample (Martino; col. 8, lines 5-46). It is readily apparent that this window of 8-32 words depends of the length of each word and/or sentence that comprises the windowed text, and as such, is a form of "deviational cue".

(E) Claim 26 differs from claim 1 by reciting an article of manufacture having a memory and instructions stored upon the memory for performing the method of claim 1. As per these limitations, Martino discloses a combination of a computer program executing on a conventional data processing system having either a RAM (124), hard disk drive (126), floppy disk drive (127), or CD-ROM (132) (all read on "memory") that stores sets of instructions and data structures (151-156) for performing the steps of his invention (Martino; col. 15, lines 43-50; col. 16, lines 38-54; and fig. 6).

The remaining limitations of claim 26 are as addressed above in the rejection of claim 1, and are therefore rejected for the same reasons given above for claim 1.

6. Claims 4-5, 7-12, 15-25, and 27 are rejected under 35

U.S.C. 103(a) as being unpatentable over Martino, et al.

(5,913,185), as applied to claims 1-3 above, and further in view

of Biber ("The Multi-Dimensional Approach to Linguistic Analyses of Genre Variation: An Overview of Methodology and Findings", *Computers and the Humanities*, vol. 26, pp. 331-345, published in 1993).

(A) As per claims 4-5 and 7-12, Martino, as noted above, generally discloses generating a log of punctuation (such as periods, commas, or quotation marks), capitalization, font, or paragraphs (read on "set of nonstructural, surface cues" and "lexical cues"), since changes therein typically signify (i.e., "cue") a language/genre change in the text (Martino; col. 10, lines 46 to col. 11, line 15 and fig. 4). But Martino fails to expressly disclose the use of "string recognizable construction cues" and some of the "lexical cues" recited by Applicant. However, the cues primarily recited by Applicant are old and well known in the art, as evidenced by Biber, who discloses identifying different textual genres via the use of grammatical tagging that indicate features including: modals (reads on "modal auxiliaries"); specialized verb classes (reads on "forms of the verb 'be'"; "words ending in 'ed'"); pronouns and pro-verbs (read on "first person pronouns" and "second person pronouns"); prepositional phrases; negation (reads on "overt negatives"); time and place adverbials (read on "sentences starting with adverb and a comma"); etc. (See page 333, section 2.2 - Overview

of Computational Methods of Biber).

One having ordinary skill in that at the time of invention would have found it obvious to incorporate Biber's tags as cues within the Martino system with the motivation of performing stylistic investigations applicable across many text and text types (Biber; page 331, first two paragraphs after section 1. *Introduction*).

Moreover, insofar as Applicant recites "represents a one of" in the claims, it is irrelevant whether or not Martino and Biber collectively disclose each and every cue that is recited.

(B) Claims 15-18 repeat limitations addressed above in the rejection of claims 4-5, 7-12, and 13-14, above, and are therefore rejected for the same reasons given above for claims 4-5, 7-12, and 13-14.

Further, as noted above, insofar as Applicant recites "includes at least one ... representing a one of" in the claims, it is irrelevant whether or not Martino and Biber collectively disclose each and every cue that is recited.

(C) Claim 19 differs from claim 1 by reciting the additional step of "determining a relevancy to the text of each facet ...

~~using the cue vector and a weighting vector".~~ As per this limitation, Biber clearly discloses identifying at least fifteen

different spoken and written genres, from official documents, press reportage, etc., to telephone conversation, and relating each genre to the linguistic characteristics of texts (i.e., "cues") based on "dimension scores" or "factor scores", and wherein the text's relevancy to a particular genre is determined based on "dimensions" or "factors" (read on "facets", e.g., Informational vs. Involved; Narrative vs. Non-Narrative; Overt Expression or Persuasion") (Biber; pages 334-337, section 2.3 - *Five Basic Dimensions of Variations in English; Analysis of Spoken and Written Genres*; note especially tables 2 and 3).

The remaining limitations of claim 19, Martino's teachings with respect to weighted vectors, and the obviousness of cue vectors in light of Martino's disclosure, are all as discussed above in the rejection of claim 1, and incorporated herein. Further, the motivation for combining the respective teachings of Martino and Biber is as given above in the rejection of claims 4-5 and 7-12, and also incorporated herein.

(D) Claims 20-21 repeat limitations addressed above in the rejection of claims 2, 4, 6, 7, and 13, above, and are therefore rejected for the same reasons given above for claims 2, 4, 6, 7, and 13.

(E) As per claims 22-25, Biber clearly discloses "dimensions" or

factors" (read on "facets") such as Informational vs. Involved; Narrative vs. Non-Narrative; Overt Expression or Persuasion" and further discloses at least fifteen different spoken and written genres, including: official documents; press reportage (reads on "press report type"); press editorials (reads on "editorial opinion type"); etc., (Biber; pages 334-337, section 2.3 - *Five Basic Dimensions of Variations in English; Analysis of Spoken and Written Genres*; note especially tables 2 and 3).

Moreover, insofar as Applicant recites "includes at least a one of" in the claims, it is irrelevant whether or not Martino and Biber collectively disclose each and every facet or genre that is recited.

(F) Claim 27 differs from claim 19 by reciting an article of manufacture having a memory and instructions stored upon the memory for performing the method of claim 19. As per these limitations, Martino discloses a combination of a computer program executing on a conventional data processing system having either a RAM (124), hard disk drive (126), floppy disk drive (127), or CD-ROM (132) (all read on "memory") that stores sets of instructions and data structures (151-156) for performing the steps of his invention (Martino; col. 15, lines 43-50; col. 16, lines 38-54; and fig.--6).

The remaining limitations of claim 27 are as addressed above

Serial No: 09/100,189
Art Unit: 2747

-11-

in the rejection of claim 19, and are therefore rejected for the same reasons given above for claim 19.

Conclusion

7. The prior art made of record and not relied upon is considered pertinent to Applicant's disclosure. The cited but not applied prior art teaches a method and apparatus for text analysis (4,773,009); an information processing expert system for text analysis and predicting public opinion based information available to the public (4,930,077); processing natural language text using autonomous punctuational structure (5,111,398); a method for checking the correct and consistent use of units or chemical formulae in a text processing system (5,159,552); a system for evaluating a psychological effect of a document (5,424,945); a system for searching a corpus of document images by user specified document layout components (5,999,664); and a natural language determination using correlation between common words (6,023,670). The cited but not applied prior art of record further includes foreign patents that are Applicant's related work (JP 11-120206 A & EP 0 899 417 A2).

8. **Any response to this action should be mailed to:**

Commissioner of Patents and Trademarks
Washington, D.C. 20231

Serial No: 09/100,189
Art Unit: 2747

-12-

or faxed to:

(703) 308-6306

Or:

(703) 308-6296

For informal or draft communications, please label "PROPOSED" or "DRAFT" on the front page of the communication, and do NOT sign the communication.


Hand-delivered responses should be brought to Crystal Park II, 2021 Crystal Drive, Arlington. VA., Sixth Floor (Receptionist).

9. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Joseph Thomas, whose telephone number is (703) 305-9588. The examiner can normally be reached on Monday through Thursday from 8:30 AM to 5:00 PM. The examiner can also be reached on alternate Fridays.

If attempts to reach the examiner are unsuccessful, the examiners' supervisor, Forester W. Isen, can be reached at (703) 305-4386.

Any inquiry of a general nature or relating to the status of this application should be directed to the Group receptionist whose telephone number is (703) 305-3900.

jt
March 9, 2000


Joseph Thomas
Primary Examiner
Art Unit 2747